Osteoarthritis and Cartilage 19 (2011) 921–929

Osteoarthritis and Cartilage

Review

NSAIDs vs acetaminophen in knee and hip osteoarthritis: a systematic review regarding heterogeneity influencing the outcomes

S.P.J. Verkleij †*, P.A.J. Luijsterburg †, A.M. Bohnen †, B.W. Koes †, S.M.A. Bierma-Zeinstra †‡

† Department of General Practice, Erasmus MC, University Medical Center Rotterdam, the Netherlands
‡ Department of Orthopaedic Surgery, Erasmus MC, University Medical Center Rotterdam, the Netherlands

ARTICLE INFO

Article history:
Received 23 February 2011
Accepted 29 April 2011

Keywords:
Osteoarthritis
Heterogeneity
Meta-analysis
NSAID
Acetaminophen

SUMMARY

Objective: To identify sources of heterogeneity (statistical, methodological, and clinical) in studies evaluating non-steroidal anti-inflammatory drugs (NSAIDs) vs acetaminophen in patients with knee and hip osteoarthritis (OA) to elucidate variations in outcomes.
Method: A database search (1966 to January 2010) was made for (randomized) controlled trials ((R)CTs) comparing NSAIDs vs acetaminophen in knee and hip OA. Extracted data included baseline demographic/clinical characteristics, outcomes at follow-up, and characteristics of study design. Heterogeneity was examined with subgroup analyses by exploring changes in effect size and with I^2 of Higgins. Pain measures were expressed as standardized mean differences.
Results: 15 RCTs, including 21 comparisons of NSAIDs and acetaminophen were included. Statistical heterogeneity was absent (Cochran's Q-test = 14.11; $I^2 = 0$; $P = 0.78$). Moderate clinical heterogeneity was found for comparisons which included both hip and knee OA vs knee OA only ($I^2 = 51$; $P = 0.09$). NSAIDs seemed slightly more effective than acetaminophen if more patients with hip OA were included. However, the pooled effect sizes of comparisons with knee OA vs both knee and hip OA are equal. Low clinical heterogeneity was found for comparisons with low dosage of acetaminophen, normal dosage of NSAIDs, and moderate pain intensity at baseline. Low methodological heterogeneity was found for comparisons with a short duration.
Conclusion: Future trials should present the results of hip and knee OA separately, as moderate clinical heterogeneity was found. There might be differences in effectiveness of NSAIDs vs acetaminophen in patients with hip vs knee OA. No significant methodological and statistical heterogeneity was found in studies evaluating NSAIDs vs acetaminophen.

Introduction

Guidelines for the treatment of non-traumatic knee complaints, recommend acetaminophen as the first-choice analgesic in treating pain due to osteoarthritis (OA)[1–4]. This recommendation is based on a review showing a superior effectiveness of acetaminophen compared to placebo in treating pain due to OA[5]. In addition, recent reviews reported only small improvements (effect sizes from 0.2 to 0.37) in pain in favor of non-steroidal anti-inflammatory drugs (NSAIDs) compared to acetaminophen[5–8]. However, NSAIDs were consistently associated with substantially more side effects[9,10].

Systematic reviews integrate the results of original studies. In order to pool the data of original studies in a quantitative manner (meta-analysis), the reviews investigate whether statistical heterogeneity exists (i.e., when the statistical difference between studies is larger than expected by chance). If studies are statistically heterogeneous, there is no added value of pooling while the pooled effect size might be biased. No statistical heterogeneity was reported in the four reviews comparing the efficacy of NSAIDs vs acetaminophen[5–8].

There are two other types of heterogeneity. Clinical heterogeneity can result from differences in the characteristics of the included patients, from the interventions applied, and from the use of different outcome measures. Methodological heterogeneity arises when different study designs are used, and with differences in the degree of control over bias[11]. However, data on clinical and methodological heterogeneity in systematic reviews are scarce.

A meta-analysis investigating the efficacy of NSAIDs vs placebo in patients with knee OA found a higher effect size (0.32) for studies

* Address correspondence and reprint requests to: S.P.J. Verkleij, Department of General Practice, Erasmus MC, University Medical Center Rotterdam, P.O. Box 2040, 3000 CA Rotterdam, the Netherlands. Tel: 31-10-703-2-117; Fax: 31-10-703-2-127.
E-mail address: s.verkleij@erasmusmc.nl (S.P.J. Verkleij).

1063-4584/$ – see front matter
doi:10.1016/j.joca.2011.04.013

that required a flare of symptoms compared to studies without a flare design (effect size 0.23)[12]. Vlad *et al.* examined differences in study characteristics among glucosamine trials in OA; they found consistently higher effect sizes in trials with industry funding (0.47) vs trials without industry funding (0.05)[13]. Furthermore, one randomized controlled trial (RCT) reported that NSAIDs are superior in patients with moderate to severe pain, but in patients with mild intensity pain the differences between NSAIDs and acetaminophen were negligible[14]. Bradley *et al.*, reported that higher baseline pain intensity predicted greater pain relief at follow-up[15].

The present review assesses the presence of statistical, methodological and clinical heterogeneity of RCTs comparing NSAIDs vs acetaminophen in knee and hip OA, with the aim to elucidate variations in outcomes.

Methods

Prisma recommendations

When executing the review the recommendations of the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA)[16] were followed.

Protocol

Methods of the data-analysis and selection criteria were specified in advance and documented in a protocol (not published).

Search strategy

A search (1966 to January 2010) of PubMed, Embase, Web of Science, Cinahl, Scopus, and the Cochrane Library (Cochrane database of systematic reviews, database of abstracts of reviews of effects, and Cochrane central register of controlled trials) was performed to identify studies evaluating the effectiveness of NSAIDs vs acetaminophen in OA. The databases were searched using the following terms: "Non-steroidal anti-inflammatory agent", "Paracetamol", "Acetaminophen", and "OA". A detailed description of the full electronic search strategies is provided in Appendix A.

Selection criteria

Inclusion criteria

All RCTs, controlled clinical trials (CCTs), and quasi-RCTs (qRCTs) comparing NSAIDs and acetaminophen in OA patients aged 18 years and older were included. Trials with a cross-over design were also eligible. OA was either determined clinically and/or with radiography. Studies in the English, German, or Dutch language were included.

Exclusion criteria

Studies were excluded if they did not pertain to OA or concerned non-oral pharmacologic therapy.

Study collection

Two authors (SV and PL) independently evaluated all eligible titles, abstracts, and full-text articles based on the inclusion and exclusion criteria. Disagreement was resolved by discussion. From relevant articles the references lists were searched for additional articles.

Data extraction

One author (SV) performed the data extraction using a standardized form. In case of uncertainty a second author was consulted. The following data were collected:

(1) Demographic and clinical characteristics at baseline (age, baseline pain intensity, previous used medication, mean duration of complaints due to OA, radiographic severity, and patient preferences); (2) outcomes at follow-up (mean difference and standard deviations (SD) in pain, use of rescue medication, co-interventions, and loss to follow-up); and (3) design characteristics (flare design, localisation of OA, dosage of NSAID and acetaminophen, duration of follow-up in weeks, sample size, criteria for eligible patients, setting and recruitment, and industry funding).

Methodological quality of the studies

The methodological quality of the included studies was independently assessed by two authors (SV and PL, BK, AB, or SB) using a predefined list by the Cochrane Collaboration[17], which is based on the Delphi criteria[18]. Dissimilarity between researchers was resolved by discussion. Scored items were: (1) randomization procedure (randomization generation and randomization concealment), (2) blinding (participants, care provider, and outcome assessors), (3) incomplete outcome data (drop-out rates and number of participants analysed in the group of allocation), (4) selective outcome reporting, and (5) other sources of bias related to comparability of study groups at baseline, co-interventions, compliance to treatment, and timing of outcome assessment. Each item was rated as 'Yes' (indicating a low risk of bias), 'No' (indicating a high risk of bias), or 'Unclear' (indicating unclear or unknown risk of bias).

Types of outcome measures

Pain intensity was used as the main outcome measure, which was assessed by standardized and validated scales or questionnaires, such as the Western Ontario and McMaster Osteoarthritis index (WOMAC)[19] or the visual analog scale (VAS). If WOMAC or VAS data were not available, other measures of pain were used (e.g., pain at rest, pain during walking, etc.). If possible, all scales and questionnaires were converted to a standardized scale from 0 to 100 (0 = no pain; 100 = worst pain ever). In addition, standardized mean differences (SMD) of pain were used to measure the magnitude of the treatment effect (with negative values in favor of NSAIDs). According to Cohen, a treatment effect of 0.2–0.5 is regarded as a small effect, 0.5–0.8 is a medium effect, and a score of 0.8 or higher a large treatment effect[20]. If data on the primary outcome were missing, authors of the included studies were contacted.

Data analyses

Extracted data on number of patients, mean differences in pain intensity, and SD were used to estimate the pooled SMD. If SDs were not reported, they were obtained from confidence intervals or standard errors. If necessary, authors of the specific study were asked for more information. Otherwise, baseline SDs were used or imputations were made using SDs from a similar study. Regarding cross-over design, only data of the first comparison before crossing-over were included for analysis.

If eligible studies compared different types or multiple dosages of NSAIDs to acetaminophen, all types and dosages were included in the analyses as separate comparisons. To avoid double counting of the acetaminophen group, the number of participants in the acetaminophen group was divided by the number of comparisons of NSAIDs.

Heterogeneity was assessed using subgroup analyses. A statistical test to indicate the extent of heterogeneity is the Cochran's Q-test. However, this test has relatively low power to detect heterogeneity when meta-analysis includes a small number of studies. Therefore, the degree of heterogeneity was quantified with the method of Higgins *et al.*[21,22]. With the method of Higgins *et al.*, the variation across a study caused by heterogeneity and not by chance is measured with I^2. I^2 is calculated with the Cochran's Q-test and the degrees of freedom. I^2 values range from 0% to 100%: an I^2 of 0% indicates no observed heterogeneity, and 25%, 50%, and 75% indicate low, moderate, and high heterogeneity, respectively. An I^2 of 50% means that half of the total variability among effect sizes is not caused by sampling error, but by true heterogeneity between studies[21]. Besides using I^2, the extent of heterogeneity was also assessed by exploring the changes in SMD.

We hypothesized that the following trial characteristics might influence the results: baseline pain intensity, flare design, use of previous pain medication, radiographic severity, localisation of OA, methodological quality, duration of follow-up, sample size, dosage of acetaminophen, types and dosage of NSAID, and industry funding.

To assess the possibility of publication bias we evaluated a funnel plot visually for symmetry.

Baseline pain intensity was categorized as 'moderate' (≤55 on a scale of 0–100) vs 'high' (>55). Dosage of NSAID was categorized, based on the Dutch information leaflet, as 'normal' for ibuprofen 1200 mg, rofecoxib 12.5 mg, naproxen 440 and 660 mg, and floctafenin 800 mg. Dosage of NSAID was categorized as 'high' for ibuprofen 2400 mg, diclofenac 150 mg, rofecoxib 25 mg, aceclofenac 200 mg, celecoxib 200 mg, and naproxen 750 mg. Types of NSAIDs were categorized as cyclo-oxygenase-2 inhibitors (rofecoxib and celecoxib), phenylacetic acids (diclofenac and aceclofenac), or propanoic-phenolic acids (ibuprofen and naproxen). Dosage of acetaminophen was categorized as 'low' (<3000 mg) or 'high' (≥3000 mg). Industry funding was categorized as 'absent' (no reports of industry funding were reported in the study) and 'present'. Duration of follow-up was defined as 'short follow-up' (<6 weeks) or 'long follow-up' (≥6 weeks). Sample size was divided into 'small' (<100 patients) and 'large' (>100 patients).

We used both fixed-effect and random-effect models to analyse the results. If both models presented the results equally, only the results of the fixed-effect model are presented. Review manager 5.0 and STATA 11 (StataCorp. 2009. Stata Statistical Software: Release 11. College Station, TX: StataCorp LP) were used for the analyses.

Results

The literature search yielded 1659 potentially eligible studies. Finally, 14 articles were included in the present review (Fig. 1)[23–36]. One article published data from two separate studies[30]. Four studies had more than one comparison (range 2–3)[25,27,32,33]. In total, 15 studies and 21 comparisons were included in the present review.

Study characteristics

Table I presents the characteristics of the included studies. Both fixed-effect and random-effect models yielded the same results, so we presented the results of the fixed-effect analysis only. A total of 5133 patients were randomized, of whom 3275 received an NSAID and 1858 received acetaminophen. All studies were RCTs. Three studies had a cross-over design[14,29,30]. All studies included patients with knee OA. Six studies also included patients with hip OA[24,29–31,35]. Seven different NSAIDs (naproxen, celecoxib, rofecoxib, floctafenin, ibuprofen, diclofenac, and aceclofenac) were compared to acetaminophen. With the exception of one study[36], all studies used a wash-out period prior to randomization. Three studies also required a flare of symptoms after the wash-out period[32,33,35]. Industry funding was reported in 11 of the 15 studies[23,24,27,28,30–35]. All studies reported hip or knee pain intensity as a primary outcome for effectiveness of NSAID vs acetaminophen. Pain intensity was measured by the WOMAC pain scale, a 0–100 mm VAS, or a 4/5-point numerical scale (pain at rest). The mean duration of follow-up was 15.4 (range 1–104) weeks.

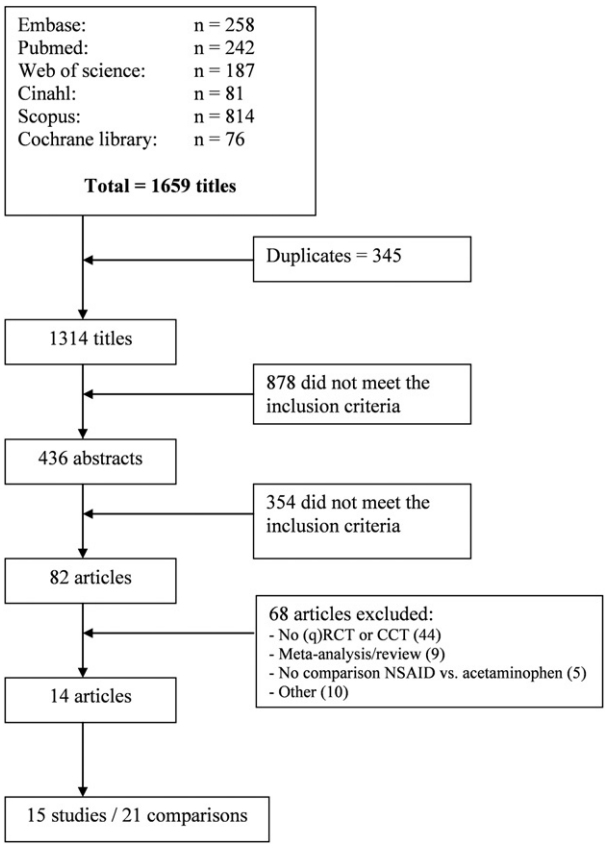

Fig. 1. Flowchart of the study selection.

Table II presents the risk of bias assessment. With the exception of one study[34], all studies were blinded. The procedure of randomization generation, randomization concealment, and blinding was satisfactorily reported in only four studies[23,29,35,36].

Effectiveness of NSAID vs acetaminophen

A total of 14 studies and 20 comparisons provided analyzable data of 2991 patients in the NSAID group vs 1561 patients in the acetaminophen group. The pooled SMD was −0.29 [95% confidence interval (95% CI): −0.35 to −0.22], referring to a small treatment effect in favor of NSAIDs (Fig. 2). This finding is in accordance with results from previous reviews[5–8].

Heterogeneity

Statistical heterogeneity

No statistical heterogeneity was found between the included comparisons evaluating NSAIDs vs acetaminophen in knee and hip OA (Cochran's Q-test = 14.11; $I^2 = 0$; $P = 0.78$) (Fig. 2).

Table I
Characteristics of studies evaluating the effectiveness of NSAIDs vs acetaminophen in patients with knee and hip OA

Study	n	Joints (%)	Mean age (years)	NSAID and dosage (mg/day)	Acet# dosage (mg/day)	Baseline pain NSAID	Baseline pain Acet	Duration (weeks)	Flare design	Industry involvement
Battle-Gualda *et al.*[23]	168	Knee	62.4	Aceclofenac 200	3000	62.2¶	62.4¶	6	Wash out	Yes
Boureau *et al.*[24]	222	Hip (29.5) Knee (70.5)	66.5	Ibuprofen 1200	3000	71.3†	72.2†	2	Wash out	Yes
Bradley *et al.*[25]	184	Knee	56.5	Ibuprofen 1200 Ibuprofen 2400	4000	50§ 49§	54§	4	Wash out	No
Case *et al.*[26]	54	Knee	62.5	Diclofenac 150	4000	37†	31.8†	12	Wash out	No
Geba *et al.*[27]	382	Knee	62.6	Celecoxib 200 Rofecoxib 12.5 Rofecoxib 25	4000	Unknown†	Unknown†	6	Wash out	Yes
Golden *et al.*[28]	310	Knee	60.7	Naproxen 440/660	4000	35.3§	34.5§	1	Wash out	Yes
Lequesne *et al.*[29],†	192	Hip (33.5) Knee (66.5)	64.7	Floctafenin 800	3000	64.5¶	60.9¶	1.7	Wash out	No
Pincus *et al.*[14],†	227	Hip (22) Knee (78)	61.5	Diclofenac 150	4000	40.2†	42.1†	9	Wash out	Yes
Pincus *et al.*[30],†	235	Hip (16.4) Knee (83.6)	63.5	Celecoxib 200	4000	48.6†	52.8†	6	Wash out	Yes
	250	Hip (15.5) Knee (84.5)	63.7	Celecoxib 200	4000	52†	51.6†	6	Wash out	Yes
Schnitzer *et al.*[33]	1578	Knee	62.1	Celecoxib 200 Rofecoxib 12.5 Rofecoxib 25	4000	Unknown†	Unknown†	6	Yes	Yes
Schnitzer *et al.*[32]	403	Knee	59.8‡	Rofecoxib 12.5 Rofecoxib 25	1300	57.2† 62.1†	58.1†	4	Yes	Yes
Shen *et al.*[34]	20	Knee	Unknown	Rofecoxib 25	4000	68.5†	90.5†	12	Wash out	Yes
Temple *et al.*[35]	571	Hip/knee	59.3	Naproxen 750	4000	Unknown	Unknown	52	Yes	Yes
Williams *et al.*[36]	178	Knee	59.6	Naproxen 750	2600	26§	29§	104*	No	No

* Intention to treat analysis after 42 days.
† Cross-over design.
‡ Characteristics of the intention to treat analysis.
§ Rest pain measured on a four-point or a five-point scale.
† Pain measured with the WOMAC.
¶ Pain measured with a VAS.
# Acet: acetaminophen.

Table II
Risk of bias assessment of included RCTs

Study	Sequence generation	Allocation concealment	Blinding of participants	Blinding of personnel	Blinding of outcome assessors	Drop out	Participant analysed in group of allocation	Selective outcome reporting	Similar groups at baseline	Co-interventions	Compliance	Timing of outcome assessment
Battle-Gualda *et al.*[23]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Boureau *et al.*[24]	Unsure	Unsure	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Unsure	Unsure	Yes
Bradley *et al.*[25]	Unsure	Unsure	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Unsure	Yes	Yes
Case *et al.*[26]	Unsure	Unsure	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Geba *et al.*[27]	Yes	Unsure	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Unsure	Unsure	Yes
Golden *et al.*[28]	Unsure	Unsure	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes
Lequesne *et al.*[29]	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Unsure	Yes
Pincus *et al.*[14]	Unsure	Yes	Yes	Yes	Yes	Yes	Unsure	Yes	Yes	Unsure	Unsure	Yes
Pincus *et al.*[30]	Yes	Unsure	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Unsure	Unsure	Yes
	Yes	Unsure	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Unsure	Unsure	Yes
Schnitzer *et al.*[33]	Yes	Unsure	Yes	Yes	Yes	Yes	Yes	Yes	Unsure	Unsure	Yes	Yes
Schnitzer *et al.*[32]	Unsure	Unsure	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Unsure	Yes	Yes
Shen *et al.*[34]	Unsure	Unsure	Unsure	Unsure	Unsure	Yes	Unsure	Yes	Unsure	Yes	Unsure	Yes
Temple *et al.*[35]	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Unsure	Unsure	Yes	Yes
Williams *et al.*[36]	Yes	Yes	Yes	Unsure	Yes	Yes	Yes	Yes	Yes	Yes	Unsure	Yes

Methodological heterogeneity

Subgroup analysis (Table III) showed no differences in effect between comparisons with and without a flare design (SMD: −0.27; 95% CI: −0.39 to −0.16 vs −0.30; −0.38 to −0.21), between comparison that were inadequately blinded (−032; −0.60 to −0.04) vs adequately blinded comparisons (−0.29; −0.35 to −0.22), comparisons that adequately addressed with incomplete outcome data (−0.28; −0.35 to −0.21) vs comparisons that inadequately addressed incomplete outcome data (−0.32; −0.47 to −0.17), and comparisons that adequately addressed other sources of bias (−0.28; −0.45 to −0.11) vs comparisons that did not address other sources of bias adequately (−0.29; −0.36 to −0.22).

Comparisons without industry funding have a somewhat higher treatment effect in favor of NSAIDs vs comparisons with industry funding (−0.40; −0.59 to −0.22 vs −0.30; −0.38 to −0.21). The same was found for comparisons with an adequate randomization procedure (−0.39; −0.57 to −0.21) vs uncertainty in the randomization procedure (−0.27; −0.34 to −0.20), comparisons with a sample size of ≤100 patients (−0.35; −0.46 to −0.23) vs comparisons with a sample size of ≥100 patients (−0.27; −0.34 to −0.19), and for comparisons with a short follow-up (−0.35; −0.46 to −0.23) vs long follow-up (−0.26; −0.34 to −0.18). Comparisons with a short follow-up showed minor methodological heterogeneity ($I^2 = 19\%$). None of these differences were of clinical or statistical significance (Table III).

Clinical heterogeneity

Comparisons which included patients with hip and knee OA showed moderate clinical heterogeneity ($I^2 = 51\%$; $P = 0.09$) (Table III). There was a small trend for a better effectiveness of NSAIDs compared with acetaminophen in studies which included a higher percentage of patients with hip OA. However, the pooled effect sizes of the comparisons with knee OA only vs knee and hip OA are the same (Fig. 2). Comparisons with moderate baseline pain intensity, low dosage of acetaminophen, and normal dosage of NSAIDs showed only low clinical heterogeneity (I^2 of 19%, 21%, and 26%, respectively). Subgroup analysis (Table III) showed that comparisons with moderate pain intensity at baseline have a slightly higher treatment effect in favor of NSAIDs than comparisons with high pain at baseline (SMD: −0.37; 95% CI: −0.51 to −0.24 vs −0.25; −0.35 to −0.16). The same was found for comparisons with a low dosage of acetaminophen (−0.36; −0.49 to −0.24) vs a high dosage of acetaminophen (−0.26; −0.34 to −0.18), comparisons with a normal dosage of NSAIDs (−0.32; −0.43 to −0.20) vs high dosage of NSAIDs (−0.27; −0.35 to −0.19), and for comparisons of propanoic-phenolic acids (−0.35; −0.48 to −0.21) vs phenylacetic acids and coxibs (−0.31; −0.49 to −0.12 and −0.24; −0.33 to −0.16, respectively).

Publication bias

Based on the funnel plot, there appears to be no indication for publication bias (Fig. 3).

Discussion

This systematic review investigated various sources of heterogeneity (statistical, clinical, and methodological) of RCTs evaluating the effectiveness of acetaminophen vs NSAIDs in patients with knee and hip OA in relation to pain outcomes. We found moderate clinical heterogeneity for the five studies that included both hip and knee OA[14,24,29,30,35]. Unfortunately, data on knee and hip OA were not presented separately in the available studies. Our forest plot showed a higher effect size in favor of NSAIDs in studies that included a higher percentage of patients with hip OA. Svensson *et al.* also found differences in effectiveness in patients with hip and knee OA treated with naproxen. Yet, they found a greater

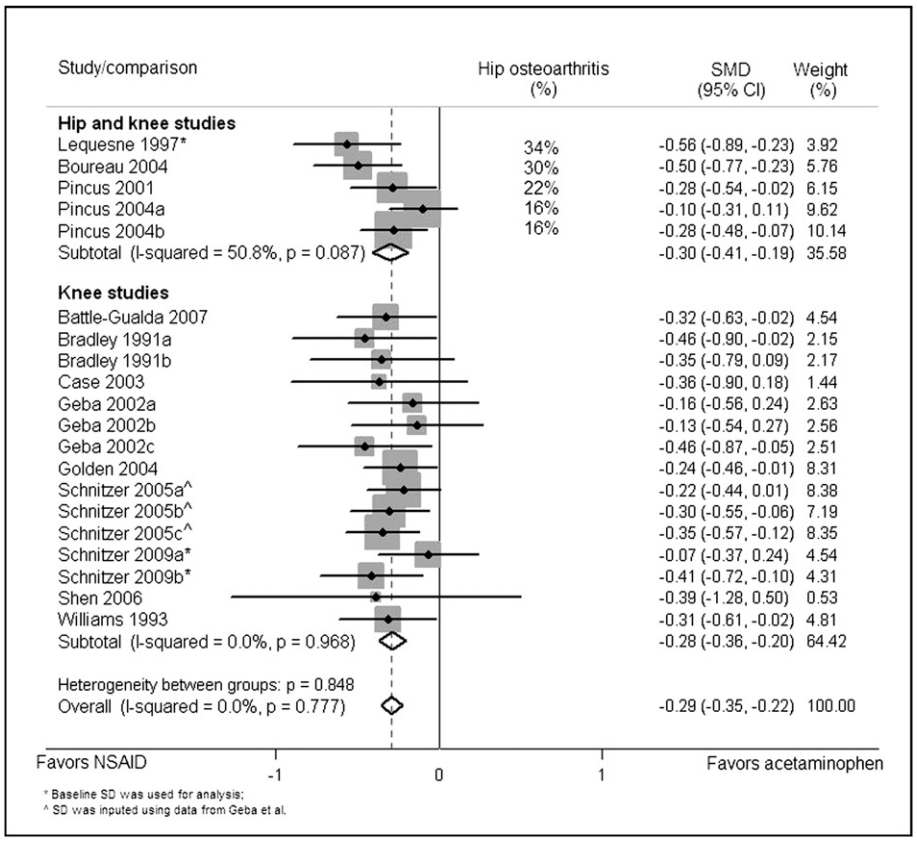

Fig. 2. Forest plot of SMD for pain improvement in NSAIDs vs acetaminophen stratified by location of OA.

improvement in patients with knee OA compared with hip OA[37]. Future research should stratify the results of hip and knee OA. Furthermore, our review showed low but non-significant clinical and methodological heterogeneity for comparisons evaluating low dosage of acetaminophen, normal dosage of NSAIDs, moderate pain intensity at baseline, and follow-up of $\leq$6 weeks.

Pincus *et al.* reported that efficacy of NSAIDs and acetaminophen is probably the same in patients with mild OA[14]. Our subgroup analyses showed a higher effect size in trials with moderate pain intensity at baseline compared to trials with high pain at baseline; however, the differences were small and therefore not important.

In contrast to Vlad *et al.*[13] our heterogeneity analyses showed that studies without industry funding have a slightly higher but non-significant effect size compared with studies with industry funding (SMD: −0.40; 95% CI: −0.59 to −0.22 vs −0.30; −0.38 to −0.21). This finding is in accordance with Lee *et al.*[6] In a *post hoc* stratification we looked more deeply for influences of industry funding. We stratified funded studies according to their SMD. Studies that were significantly in favor of NSAIDs showed no heterogeneity but, not surprisingly, showed a higher overall SMD (−0.33; −0.42 to −0.24) in favor of NSAIDs compared to those studies that were not significantly in favor of NSAIDs (−0.14; −0.26 to −0.02).

Except for one study[36], all included studies used a pre-treatment wash-out period before randomization. Additionally, three studies also required a flare of symptoms after medication discontinuation, before the start of the study[32,33,35]. The study of Scott-Lennox *et al.* examined the impact of flare designs on trial results and reported a more profound pain reduction in patients with an intense flare prior to treatment[38].

Four of the included studies compared rofecoxib (Vioxx) with acetaminophen[27,32–34], whereas rofecoxib was withdrawn from the market in 2004. We have included these studies in the present review because we were interested in the heterogeneity of studies evaluating the effect of NSAIDs vs acetaminophen on pain control. However, additional analysis showed that the exclusion of the rofecoxib trials did not alter the results (data not shown).

Similar to Bjordal *et al.*[12], we found that mean age of the participants was relatively low (61.6 years; SD 2.46) for patients with OA. Three studies excluded participants above 75 years of age[23,26,35] and one study excluded patients aged 85 years and older[24]. Although, prescribing NSAIDs in older adults is not without risks[39], the exclusion of these patients could cause selection bias. Therefore, future studies should not exclude older patients but carefully screen them before inclusion and monitor them after inclusion.

In line with previous reviews[5–8], we found a significant improvement in pain in favor of NSAIDs in patients with knee and/or hip OA. We included three additional studies[23,32,35] that were published since the last review appeared in 2006[5].

The effectiveness of acetaminophen vs NSAIDs is an important question in all guidelines for OA. It is striking, that there is a relative lack of good quality randomized studies that evaluate the effectiveness of acetaminophen vs NSAIDs. Therefore, high quality research is needed to substantiate the effectiveness of NSAIDs over acetaminophen.

Limitations

The present review has some limitations. First, although no publication bias was revealed (Fig. 3), we cannot be certain that all published/unpublished studies were retrieved. Secondly, four studies included more than one comparison of NSAIDs[25,27,32,33]. We

Table III
Pooled effect estimates of clinical and methodological heterogeneity

	Number of studies/ comparisons	SMD (95% CI)	I^2
All studies	14/20	−0.29 (−0.35 to −0.22)	0
Methodological characteristics			
Flare design			
Present	2/5	−0.27 (−0.39 to −0.16)	0
Absent	12/15	−0.30 (−0.38 to −0.22)	0
Industry funding*			
Present	10 studies	−0.30 (−0.38 to −0.21)	0
Absent	4 studies	−0.40 (−0.59 to −0.22)	0
Randomization procedure			
Adequate	3/3	−0.39 (−0.57 to −0.21)	0
No/unsure	11/17	−0.27 (−0.34 to −0.20)	0
Blinding			
Adequate	12/18	−0.29 (−0.35 to −0.22)	0
No/unsure	2/2	−0.32 (−0.60 to −0.04)	0
Incomplete outcome data			
Adequate	10/16	−0.28 (−0.35 to −0.21)	0
No/unsure	4/4	−0.32 (−0.47 to −0.17)	0
Bias			
Adequate	3/3	−0.28 (−0.45 to −0.11)	0
No/unsure	11/17	−0.29 (−0.36 to −0.22)	0
Sample size			
<100 in NSAID group	7/10	−0.35 (−0.47 to −0.23)	0
>100 in NSAID group	7/10	−0.27 (−0.34 to −0.19)	0
Duration of follow-up			
<6 Weeks	5/7	−0.35 (−0.46 to −0.23)	19%
≥6 Weeks	9/13	−0.26 (−0.34 to −0.18)	0
Clinical characteristics			
Localisation of OA			
Knee	9/15	−028 (−0.36 to −0.20)	0
Knee and hip	5/5	−0.30 (−0.41 to −0.19)	***51%***
Baseline pain intensity*			
Moderate	7/8	−0.37 (−0.51 to −0.24)	19%
High	5/6	−0.25 (−0.35 to −0.16)	0
Missing	2/6		
Dosage of acetaminophen†			
Low	2/3	−0.36 (−0.49 to −0.24)	21%
High	12/17	−0.26 (−0.34 to −0.18)	0
Dosage of NSAID‡			
Normal	7 comparisons	−0.32 (−0.43 to −0.20)	26%
High	13 comparisons	−0.27 (−0.35 to −0.19)	0
Type of NSAID§			
Coxib	6/11	−0.24 (−0.33 to −0.16)	0
Phenylacetic acids	3/3	−0.31 (−0.49 to −0.12)	0
Propanoic-phenolic acids	4/5	−0.35 (−0.48 to −0.21)	0
Other	1/1		

Bold = $P < 0.10$.
I^2 = measure of heterogeneity; SMD (negative values are in favor of NSAIDs).
* With multi-group studies, only the highest dosage was included for analysis.
† Medication acetaminophen: low ≤2600 mg.
‡ Medication NSAID: normal: ibuprofen 1200 mg, rofecoxib 12.5 mg, naproxen 440/660 mg, and floctafenin 800 mg; medication NSAID high: ibuprofen 2400 mg, diclofenac 150 mg, rofecoxib 25 mg, naproxen 750 mg, aceclofenac 200 mg, and celecoxib 200 mg.
§ Cyclo-oxygenase-2 inhibitors (coxibs) are rofecoxib and celecoxib; phenylacetic acids are diclofenac and aceclofenac; propanoic-phenolic acids are ibuprofen and naproxen; other: floctafenin.
* High baseline pain was defined as a pain score of 55 or higher on a scale of 0–100; moderate pain intensity was a pain score of 55 or lower.

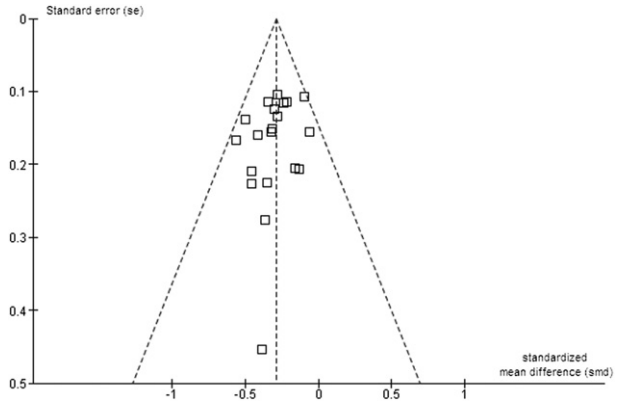

Fig. 3. Funnel plot for publication bias.

analysed each comparison separately and divided the number of participants of the acetaminophen group by the number of comparisons of NSAIDs which could have biased the results, possibly leading to an underestimation. Thirdly, one of our aims was to examine relevant trial characteristics that may cause heterogeneity. However, not all clinical features were always reported satisfactorily. For example, data on the use of previous pain medication was only reported satisfactorily in four trials (which included nine comparisons)[26,27,30,33], the mean duration of complaints was reported in four trials (which included five comparisons)[23,24,30,36], and radiographic severity was only reported in five trials (which included six comparisons)[14,23,26,30,36]. These characteristics may have influenced the reported results. Moreover, Case *et al.* reported that prior use of NSAIDs predicted a better response of NSAIDs compared to acetaminophen[26].

Another limitation is that, due to the small numbers we performed subgroup analysis with only one study characteristic (univariable analysis). Based on univariable analyses it is impossible to draw broad conclusion. Furthermore, it was not possible to study interaction effects of the treatment in subgroups of patients. Therefore, future meta-analyses should focus on individual patient data (IPD). The use of IPD in meta-analyses has been described as the gold standard[40], allowing to assess the existence of heterogeneity more reliably. Furthermore, IPD can be used to investigate specific treatment effects in various subgroups[41].

Conclusion

In conclusion, future trials should present the results of patients with hip and knee OA separately, as we found moderate heterogeneity in trials that included patients with both knee and hip OA. Furthermore, no clinically relevant statistical, methodological, or clinical heterogeneity was found for studies evaluating the effectiveness of NSAIDs vs acetaminophen in knee and hip OA.

Authors' contributions

All authors made substantial contributions to the conceptualization, design, data collection, interpretation, drafting, and revisions and approved the final version.

Conflict of interest

The authors declare that they have no conflicts of interest.

Acknowledgment

No funding was obtained for this study.

Appendix A

Medical Subject Headings and text word search strategies performed in Embase, Cinahl, Cochrane Library, Scopus, PubMed, and Web of Science (1966 to January 2010)

Embase/Cinahl/Cochrane Library/Scopus	PubMed/Web of Science
1. Non-steroidal anti-inflammatory agent/syn*	1. Anti-inflammatory agents, Non-steroidal[mesh]
2. Paracetamol/syn*‡	2. NSAIDs[tw]
3. Acetaminophen	3. Non-steroidal anti-inflammatory agents[tw]
4. OA/syn*	4. Non-steroidal anti-inflammatory agents[tw]
5. Controlled clinical trial/lim†‡	5. Non-steroidal anti-inflammatory agents[tw]
6. RCT/lim†‡	6. Non-steroidal anti-inflammatory agents[tw]
7. Adult/lim†‡	7. Non-steroidal anti-inflammatory agents[tw]
8. Aged/lim†‡	8. Non-steroidal antirheumatic agents[tw]
9. Title–abstract–keyword§	9. Aspirin-like agents[tw]
	10. Aspirin-like agents[tw]
	11. Analgesics, anti-inflammatory[tw]
	12. Analgesics, anti-inflammatory[tw]
	13. Anti-inflammatory analgesics[tw]
	14. OR 1–13
	15. Acetaminophen[mesh]
	16. Acetaminophen[tw]
	17. Hydroxyacetanilide[tw]
	18. APAP[tw]
	19. N-acetyl-p-aminophenol[tw]
	20. p-Acetamidophenol[tw]
	21. p-Hydroxyacetanilide[tw]
	22. Paracetamol[tw]
	23. Acetamidophenol[tw]
	24. Acephen[tw]
	25. Tylenol[tw]
	26. Panadol[tw]
	27. Acamol[tw]
	28. OR 15–27
	29. OA[mesh]
	30. OA[tw]
	31. Osteoarthrosis[tw]
	32. OR 29–31

mesh = Medical subject heading; tw = text word.
* syn = Synonym. The term synonym is only used in Embase.
† lim = Limitation.
‡ Search strategy only used in Embase.
§ Search strategy only used in Scopus.

References

1. Belo JN, Bierma-Zeinstra SMA, Raaijmakers AJ, van der Wissel F. NHG-Standaard niet-traumatische knieproblemen bij volwassenen. Huisarts Wetensch 2008;51:229–40.
2. Jordan KM, Arden NK, Doherty M, Bannwarth B, Bijlsma JW, Dieppe P, *et al*. EULAR recommendations 2003: an evidence based approach to the management of knee osteoarthritis: report of a task force of the Standing Committee for International Clinical Studies Including Therapeutic Trials (ESCISIT). Ann Rheum Dis 2003;62:1145–55.
3. Zhang W, Moskowitz RW, Nuki G, Abramson S, Altman RD, Arden N, *et al*. OARSI recommendations for the management of hip and knee osteoarthritis, part I: critical appraisal of existing treatment guidelines and systematic review of current research evidence. Osteoarthritis Cartilage 2007;15:981–1000.
4. Recommendations for the medical management of osteoarthritis of the hip and knee: 2000 update. American College of Rheumatology Subcommittee on Osteoarthritis Guidelines. Arthritis Rheum 2000;43:1905–15.
5. Towheed TE, Maxwell L, Judd MG, Catton M, Hochberg MC, Wells G. Acetaminophen for osteoarthritis. Cochrane Database Syst Rev 2006:CD004257.
6. Lee C, Straus WL, Balshaw R, Barlas S, Vogel S, Schnitzer TJ. A comparison of the efficacy and safety of nonsteroidal antiinflammatory agents versus acetaminophen in the treatment of osteoarthritis: a meta-analysis. Arthritis Rheum 2004;51:746–54.
7. Zhang W, Jones A, Doherty M. Does paracetamol (acetaminophen) reduce the pain of osteoarthritis? A meta-analysis of randomised controlled trials. Ann Rheum Dis 2004;63:901–7.
8. Wegman A, van der Windt D, van Tulder M, Stalman W, de Vries T. Nonsteroidal antiinflammatory drugs or acetaminophen for osteoarthritis of the hip or knee? A systematic review of evidence and guidelines. J Rheumatol 2004;31:344–54.
9. Laine L, Smith R, Min K, Chen C, Dubois RW. Systematic review: the lower gastrointestinal adverse effects of non-steroidal anti-inflammatory drugs. Aliment Pharmacol Ther 2006;24:751–67.
10. Ong CK, Lirk P, Tan CH, Seymour RA. An evidence-based update on nonsteroidal anti-inflammatory drugs. Clin Med Res 2007;5:19–34.
11. Higgins J, Thompson S, Deeks J, Altman D. Statistical heterogeneity in systematic reviews of clinical trials: a critical appraisal of guidelines and practice. J Health Serv Res Policy 2002;7:51–61.
12. Bjordal JM, Ljunggren AE, Klovning A, Slordal L. Non-steroidal anti-inflammatory drugs, including cyclo-oxygenase-2 inhibitors, in osteoarthritic knee pain: meta-analysis of randomised placebo controlled trials. BMJ 2004;329:1317.
13. Vlad SC, LaValley MP, McAlindon TE, Felson DT. Glucosamine for pain in osteoarthritis: why do trial results differ? Arthritis Rheum 2007;56:2267–77.
14. Pincus T, Koch GG, Sokka T, Lefkowith J, Wolfe F, Jordan JM, *et al*. A randomized, double-blind, crossover clinical trial of diclofenac plus misoprostol versus acetaminophen in patients with osteoarthritis of the hip or knee. Arthritis Rheum 2001;44:1587–98.
15. Bradley JD, Katz BP, Brandt KD. Severity of knee pain does not predict a better response to an antiinflammatory dose of ibuprofen than to analgesic therapy in patients with osteoarthritis. J Rheumatol 2001;28:1073–6.
16. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gotzsche PC, Ioannidis JP, *et al*. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. J Clin Epidemiol 2009;62:e1–e34.
17. Higgins J, Altman D. Assessing risk of bias in included studies. In: Handbook for Systematic Reviews of Interventions (Version 501). The Cochrane Collaboration; 2008
18. Verhagen AP, de Vet HC, de Bie RA, Kessels AG, Boers M, Bouter LM, *et al*. The Delphi list: a criteria list for quality assessment of randomized clinical trials for conducting systematic reviews developed by Delphi consensus. J Clin Epidemiol 1998;51:1235–41.
19. Bellamy N, Buchanan WW. A preliminary evaluation of the dimensionality and clinical importance of pain and disability in osteoarthritis of the hip and knee. Clin Rheumatol 1986; 5:231–41.
20. Cohen J. Statistical Power Analysis for the Behavioral Sciences. 2nd edn. Hillsdale, NJ: Lawrence Erlbaum Assiocates; 1988
21. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ 2003;327:557–60.
22. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med 2002;21:1539–58.
23. Batlle-Gualda E, Roman Ivorra J, Martin-Mola E, Carbonell Abello J, Linares Ferrando LF, Tornero Molina J, *et al*. Aceclofenac

vs paracetamol in the management of symptomatic osteoarthritis of the knee: a double-blind 6-week randomized controlled trial. Osteoarthritis Cartilage 2007;15:900–8.
24. Boureau F, Schneid H, Zeghari N, Wall R, Bourgeois P. The IPSO study: ibuprofen, paracetamol study in osteoarthritis. A randomised comparative clinical study comparing the efficacy and safety of ibuprofen and paracetamol analgesic treatment of osteoarthritis of the knee or hip. Ann Rheum Dis 2004; 63:1028–34.
25. Bradley JD, Brandt KD, Katz BP, Kalasinski LA, Ryan SI. Comparison of an antiinflammatory dose of ibuprofen, an analgesic dose of ibuprofen, and acetaminophen in the treatment of patients with osteoarthritis of the knee. N Engl J Med 1991;325:87–91.
26. Case JP, Baliunas AJ, Block JA. Lack of efficacy of acetaminophen in treating symptomatic knee osteoarthritis: a randomized, double-blind, placebo-controlled comparison trial with diclofenac sodium. Arch Intern Med 2003;163:169–78.
27. Geba GP, Weaver AL, Polis AB, Dixon ME, Schnitzer TJ, Vioxx ACTG. Efficacy of rofecoxib, celecoxib, and acetaminophen in osteoarthritis of the knee: a randomized trial. JAMA 2002;287:64–71.
28. Golden HE, Moskowitz RW, Minic M. Analgesic efficacy and safety of nonprescription doses of naproxen sodium compared with acetaminophen in the treatment of osteoarthritis of the knee. Am J Ther 2004;11:85–94.
29. Lequesne M, Fannius J, Reginster JY, Verdickt W, du Laurier MV. Floctafenin versus acetaminophen for pain control in patients with osteoarthritis in the lower limbs. Franco-Belgian Task Force. Rev Rhum Engl Ed 1997;64:327–33.
30. Pincus T, Koch G, Lei H, Mangal B, Sokka T, Moskowitz R, *et al*. Patient Preference for Placebo, Acetaminophen (paracetamol) or Celecoxib Efficacy Studies (PACES): two randomised, double blind, placebo controlled, crossover clinical trials in patients with knee or hip osteoarthritis. Ann Rheum Dis 2004;63:931–9.
31. Pincus T. Clinical evidence for osteoarthritis as an inflammatory disease. Curr Rheumatol Rep 2001;3:524–34.
32. Schnitzer TJ, Tesser JR, Cooper KM, Altman RD. A 4-week randomized study of acetaminophen extended-release vs rofecoxib in knee osteoarthritis. Osteoarthritis Cartilage 2009; 17:1–7.
33. Schnitzer TJ, Weaver AL, Polis AB, Petruschke RA, Geba GP, Groups V-aV-S. Efficacy of rofecoxib, celecoxib, and acetaminophen in patients with osteoarthritis of the knee. A combined analysis of the VACT studies. J Rheumatol 2005;32:1093–105.
34. Shen H, Sprott H, Aeschlimann A, Gay RE, Michel BA, Gay S, *et al*. Analgesic action of acetaminophen in symptomatic osteoarthritis of the knee. Rheumatology (Oxford) 2006;45: 765–70.
35. Temple AR, Benson GD, Zinsenheim JR, Schweinle JE. Multicenter, randomized, double-blind, active-controlled, parallel-group trial of the long-term (6–12 months) safety of acetaminophen in adult patients with osteoarthritis. Clin Ther 2006;28:222–35.
36. Williams HJ, Ward JR, Egger MJ, Neuner R, Brooks RH, Clegg DO, *et al*. Comparison of naproxen and acetaminophen in a two-year study of treatment of osteoarthritis of the knee. Arthritis Rheum 1993;36:1196–206.
37. Svensson O, Malmenas M, Fajutrao L, Roos EM, Lohmander LS. Greater reduction of knee than hip pain in osteoarthritis treated with naproxen, as evaluated by WOMAC and SF-36. Ann Rheum Dis 2006;65:781–4.
38. Scott-Lennox JA, McLaughlin-Miley C, Lennox RD, Bohlig AM, Cutler BL, Yan C, *et al*. Stratification of flare intensity identifies placebo responders in a treatment efficacy trial of patients with osteoarthritis. Arthritis Rheum 2001;44:1599–607.
39. Barkin RL, Beckerman M, Blum SL, Clark FM, Koh EK, Wu DS. Should nonsteroidal anti-inflammatory drugs (NSAIDs) be prescribed to the older adult? Drugs Aging 2010;27:775–89.
40. Stewart LA, Tierney JF. To IPD or not to IPD? Advantages and disadvantages of systematic reviews using individual patient data. Eval Health Prof 2002;25:76–97.
41. Bierma-Zeinstra SMA, Verhagen AP. Osteoarthritis subpopulations and implications for clinical trial design. In press.